Science User Scenario Template (version 1.1b)

Stocks, Karen I.; Schramski, Sam; Virapongse, Arika; Kempler, Lisa (2019). EarthCube User Scenario Collection. UC San Diego Library Digital Collections.
https://doi.org/10.6075/J0WQ024C

Suggestions for usage:

- Content in the yellow shaded section may be used as a high level summary.
- These are guidelines for a conversation between tech person and domain end user, not rigid script – the interviewer can rephrase, or use follow up questions to elicit more detail as needed. Questions can be asked and answered in any order, and sections can be left blank if the answer is not known.
- Consider revisiting the *'d items near the end of the conversation in case they deserve amendment

Summary Information Section
***Use Case Name** *Give a short descriptive name for the use case to serve as a unique identifier. Consider indicating desired science outcomes in the name. 120 character limit* *Note: The interviewer can revisit this near the end of the discussion by asking, "Given the discussion we've had, would you still consider "____" a good description of this endeavor? Do we want to rephrase this?* *Example: US Atlantic fish population changes*
Contacts *A list of two to many contacts, with email, plus other contact information if available. Roles for the contacts are taken from the ISO 19115 CI_RoleCode vocabulary, and at a minimum there should be an **author** specified (which we will define as the person who wrote the use case into the template, i.e. the interviewer) and a **pointOfContact** (for questions about the content of the use case) specified. The **originator** role can be used to identify the scientist who developed the use case, if different from the pointOfContact.* Example: John Doe role: author and pointOfContact Geology Data Institute jdoe@mail.com Jane Smith The Important Atmospheric Science Center Role: originator jane@mail.com

Link to Primary documentation
if available:

Permission to make public? (Yes/No)
Permission granted by:____________________
Date permission granted: YYYY-MM-DD

***Science Objectives and Outcomes**
Briefly summarize the focus of the use case work, the main steps taken and resources used, and what the team intends to achieve with this use case.

Example: Analyzing changes over time in relative population sizes in coast U.S. Atlantic ocean regions, with a particular focus on decreasing species and species of commercial importance. Done by comparing historical fish abundance data at the US SWFSC and with commercial landings numbers and coastal use data through time, using a fisheries model.

Note: revisit this section also at the end of the discussion

Overarching Science Driver
For example, does your research align with NSF science drivers and/or EarthCube Science Strategic Plan (http://earthcube.org/document/2015/earthcube-strategic-science-plan) drivers? If So, which?

- *Sources of variability*
- *Hazards*
- *Predictions*
- *State parameters*
- *Long-term trends*

If not, what phrases could summarize the overarching driver of your current objective? And please provide an expanded description.

Examples:

- *My lab is interested in understanding the effect of the increased algae in the oceans. It's been hypothesized that this disrupts other ocean habitats; our goal is to understand which species are those primarily affected and to what degree.*
- *Data shows that the oceans are warming, and certain populations are observed to be negatively impacted by the change in their environment, some to the point of extinction or near extinction. Meanwhile, the same areas are overfished, in part as a result of the reduced populations. This research will contribute to a multi-year effort to understand which populations are in statistically significant decline, not easily accounted for by natural variations.*

Scenario Detail Section

Actors
Who (or what) are the key people and/or systems involved in the project?
(If helpful, include a table here for reference)

List actors: people or things (e.g., software, specimens) that participate in the use case. Consider both the primary actors (the ones that invoke the use case and benefit from the result) and secondary actors (all other actors that participate in the use case.)

Identify key resources such as data products, data portals, sensors, models, portals, visualization tools. Products can be derived, interpretive, proxies, etc. Interviewer should ask follow-up questions to determine critical aspects of the resources: location/geographic restrictions; current status of actor-system relationships; absence or existence of the resources; known barriers in relationships. Further technical details are captured in the technical section below.

Example:

People/Roles:

- *Oceanographers: Do the field data collection and analysis (Primary Actor)*
- *NOAA Science + IT team: Collects and uploads the historical species population data*
- *Department of Fisheries: Use the resulting analyses to determine fishing quotas, dates*
- *Fisherman: Determine fishing plans, viability*

External systems/objects:

- *Data: Historical (initial input), collected (intermediate output/input), and processed (final output)*
- *Equipment for the journey, including computer hardware*
- *Software programs for analyses*
- *Ocean, including the habitats and populations*

Preconditions
"What do I need to get going on the project"
State any assumptions about what you need to get started? Any assumptions about other systems can also be stated here, for example, weather conditions. List all preconditions, requirements, assumptions, and state changes that will prevent the use case from being executed.

Example:
The ice must have melted sufficiently for the boat to travel.
Deep ocean temperature must have been above 32F for a sufficient period that the relevant wildlife are active and visible.

Critical Existing Cyberinfrastructure
List any existing cyberinfrastructure (data repositories, software, etc.) that is necessary for this use case.

Measures of Success
List the measures of success for completion of the use case.

Example:
The experiment is considered successful if the data is collected, processed, stored and uploaded, none are corrupted, all variables are collected, and are available in time for the Fisheries to establish their policies for the following year's catch.

Early in the project, a measure of readiness is that the data analysts' training is rolled out on schedule, and they don't experience significant problems accessing the data in order to set up their comparison models.

Basic Flow
Describe the steps to be followed in doing the use case if everything works right (parts of this can also be represented with a diagram, see below). This gives any browser of the document a quick view of how the work could be carried out. Document the flow as a list, a conversation, or a story. (as much as required)
Often referred to as the primary scenario or course of events.
Error states or alternate states that might be highlighted are ***not*** *included here.*

1.
2.
3.
4.
5.

...etc.

Example:

1. *Determine target test specimens and best historical data comparison set*
2. *Design an analysis protocol*
3. *Plan a trip, select a crew, organize equipment*
4. *Create data collection software templates and in-board real-time or near-real-time analysis protocols to enable early quick-check of samples and viability while on-board*
5. *Load ocean floor mapping and diver tracking software*
6. *Shove off and perform planned analyses*
7. *Return*
8. *Download historical data*
9. *Perform analyses using analysis protocol*
10. *Write up results*
11. *Upload data*
12. *Contact relevant parties.*

Alternate Flow
List any alternate flows that might occur. May include flows that involve error conditions. Or flows that fall outside of the basic flow.

Example alternate flow
1. The thaw is later so only already trained scientists and analysts are included.
2. The crew that ships out is smaller
3. Fewer habitats are observed
4. Less data is returned and processed
5. Conclusions must be drawn from a representative sample, but not all desired populations
6. Data is uploaded to the relevant location and formatted for use by the general public and the fisheries.
7. Fisheries make their decisions, as in the standard plan, on the same schedule, with the same two weeks of leeway, should they need it.
8. Their decision is published.

Activity Diagram
Draw a picture or flow chart that captures
- *Major workflow steps*
- *Actors in each step*
- *Inputs to each step and to the whole system*
- *Outputs to each step*
- *Alternate paths that are not uncommon (e.g., equipment malfunction requires a repeat of the collection step)*

Major Outcome and Post Conditions
Here we give any conditions that will be true of the state of the system after the use case has been completed.

Major Outcome *(in addition to peer-review publication):*

What happens with data and other outcomes after the project finishes: (please define)

Example Major Outcome:
All of the data has been collected for the species, has been processed and uploaded, and is available for use by the various anticipated and unanticipated users.

Example Post Conditions:
- *The data for the previous period has been removed from the server and archived. Users can only access only the new data, unless they submit a formal request.*
- *Scientists and NOAA are now more concerned about the state of the habitats in the Atlantic, as a result of the populations decline and the observed increases in ocean temperature. More proposals are expected to be drafted to do further research and study mitigation options.*

- *The ship has been taken out of the water for the winter months. No new readings will be collected until spring.*

Problems/Challenges

Describe any significant or disruptive problems or challenges that prevent or interfere with the successful completion of the activity. For each one, list

- *The challenge*
  - Why it's disruptive, harmful, or costly?
  - Who is impacted?
- *What, if any, efforts have been undertaken to fix these problems?*
  - When?
  - What was the result? Why did the fix fail?
  - Are any solutions currently being worked on?
- *What recommendations do you have for tackling this problem?*
- *How can EarthCube and/or the larger Geosciences community address this problem?*

Example:

Problem 1: Cold weather conditions and short-time frames

- *Our window for doing the research is narrow, due to*
  - *needing to wait for NOAA to supply last year's processed measurements*
  - *The short time window when the current equipment can collect samples*

- *This is a problem because we only learn an incremental amount each year to contribute to our larger understanding. Thus, it takes years to develop a full picture. We could learn faster with more data collection, faster.*

- *Potential solutions:*
  - *We could collect samples in winter if*
  1) *we had more deep water equipment that functioned in cold conditions and*
  2) *we had the software required to translate that winter habitat attributes to spring/summer conditions, such that we could extrapolate a summer fishing scenario from the winter data. This includes robust laptops and drives that can withstand cold temperatures and are well protected against water.*

- *We asked for funding for the more advanced equipment once. At the time, our proposal was denied because the oceans had not yet warmed considerably, and so it was viewed as an interesting but not practically urgent project. The bias along those lines has changed. We would need to staff up our lab to process the quantities of data we'd ideally like to collect.*

- *Can EarthCube fund the cold-water equipment?*

Problem 2: Season variations

- *Because the elements do not operate on a strict calendar, seasonal variability can make it very challenging to produce an accurate analysis of the dynamics. Other factors -- El Nino, etc. -- add more variability and outliers.*
    - *Using historical data and comparisons to other locales, we'll mitigate these challenges.*
    - *There is no specific request for assistance at this time.*

References

Provide links to other relevant information such as background, clarifying & otherwise useful source material for someone wanting a deeper understanding of this use case. Include web site links, other related project names, overall charters, additional points of contact, etc. This section is distinct from Documentation, which would be just what's needed to describe the particular use case. References and Documentation sections should be linked to each other to help consistency and completeness.

Example:
(Note: In a real use case, these bullets would each link to the source info)
- *NOAA data used as historical comparison*
- *Past ocean habitat studies about algae effects conducted by this lab*
- *Processed data results*
- *2010 Proposal for more cold-water resources (rejected)*

Notes

Any additional important information

Example:
- *We are in discussions with WHOI and Cornell Bioacoustics, as they each have well-established programs. Although we have consulted with them, we did not actually collaborate formally as part of the research.*

Technical Section

Data is often a primary actor in many science use cases. Please describe the attributes of any data being used or created as part of the use case. If you don't know, please specify the data source.

Identify sensors, portals, and final and intermediate data products that your project generates as outputs.

Data Characteristics:

- **Data Source**
  - *Example:*
    - *Historical input data is supplied by NOAA on their publicly available data cloud.*

- **Data Format**
  - *Example:*
    - *netCDF, .csv, etc.*

- **Volume (size)**
  - *Examples:*
    - *DES: 4PB, ZTF: 1PB/yr, LSST: 7PB/yr, Simulations > 10PB in 2017*

- **Velocity (e.g., real time)**
  - *Example:*
    - *LSST: 20TB/day*

- **Variety (multiple datasets, mashup)**
  - *Examples:*
    - *1) Raw Data from sky surveys*
    - *2) Processed Image data*
    - *3) Simulation data,*
    - *4) sequence data*

- **Variability**
  - *Example:*
    - *Observations are taken nightly; supporting simulations are run throughout the year, but data can be produced sporadically depending on access to*

- **Veracity/Data Quality (accuracy, precision)**
  - *Example:*
    - *Hydrographic data uses the World Ocean Circulation Experiment (WOCE) quality assessment convention and flags.*

- **Data Types**
  - *Example:*

- *Image data from observations must be reduced and compared with physical quantities derived from simulations. Simulated sky maps must be produced to match observational formats.*

Standards

List any standards that were followed for the cyberinfrastructure resources, even if already mentioned above. Standards most commonly apply to data, but can apply to models, metadata, etc.

Example:

- *netCDF data format*
- *World Ocean Circulation Experiment (WOCE) quality assessment convention and flags*

Data Visualization and Analytics

Format for visualization

Example:

- *.vtk, .tiff, .kml, netCDF*

Software

For any important software used, describe the the important characteristics (source, language, input format, output format, CPU requirements etc)

Metadata

Provide a link to, or include any relevant metadata which can add additional detail and context the dataset(s) described above.